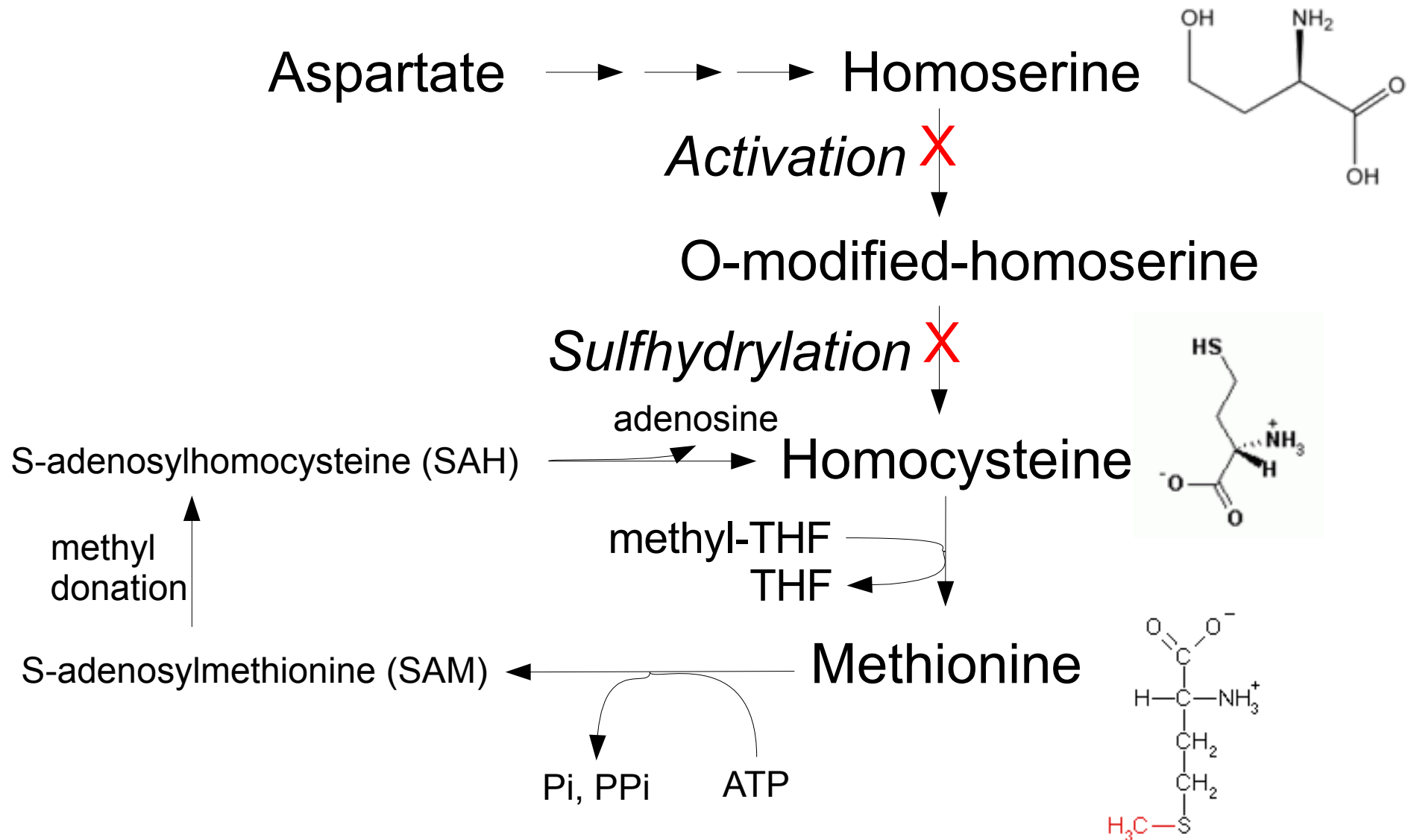


2 Topics in Comparative Genomics

Methionine synthesis
Rapid evolution of gene regulation

Morgan Price
Arkin lab
November 2007

Methionine Synthesis in *Desulfovibrios*

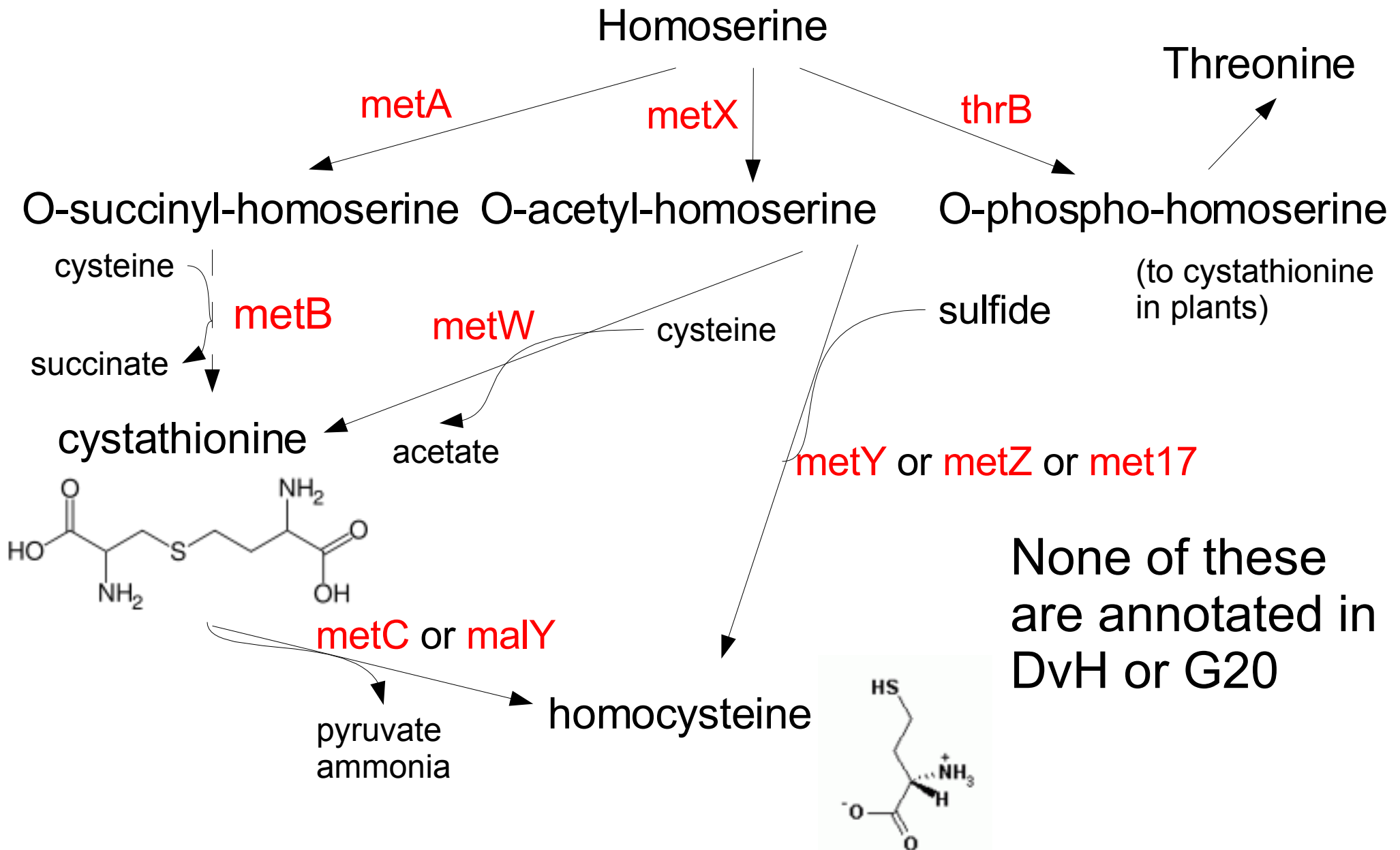


but we know DvH & G20 can do it














Hints from Experiments

- Intermediate metabolites are present
 - homoserine
 - O-succinyl-homoserine
 - cannot rule out O-acetyl- or O-phospho-homoserine
 - cystathionine
 - homocysteine
- Met labeling is as expected
- Transcriptional response to methionine
 - most effects seem secondary (cells less stressed?)
 - methionine synthase (metE) is 2x down
 - pathway flux mostly to SAM (methyl donor, polyamines)

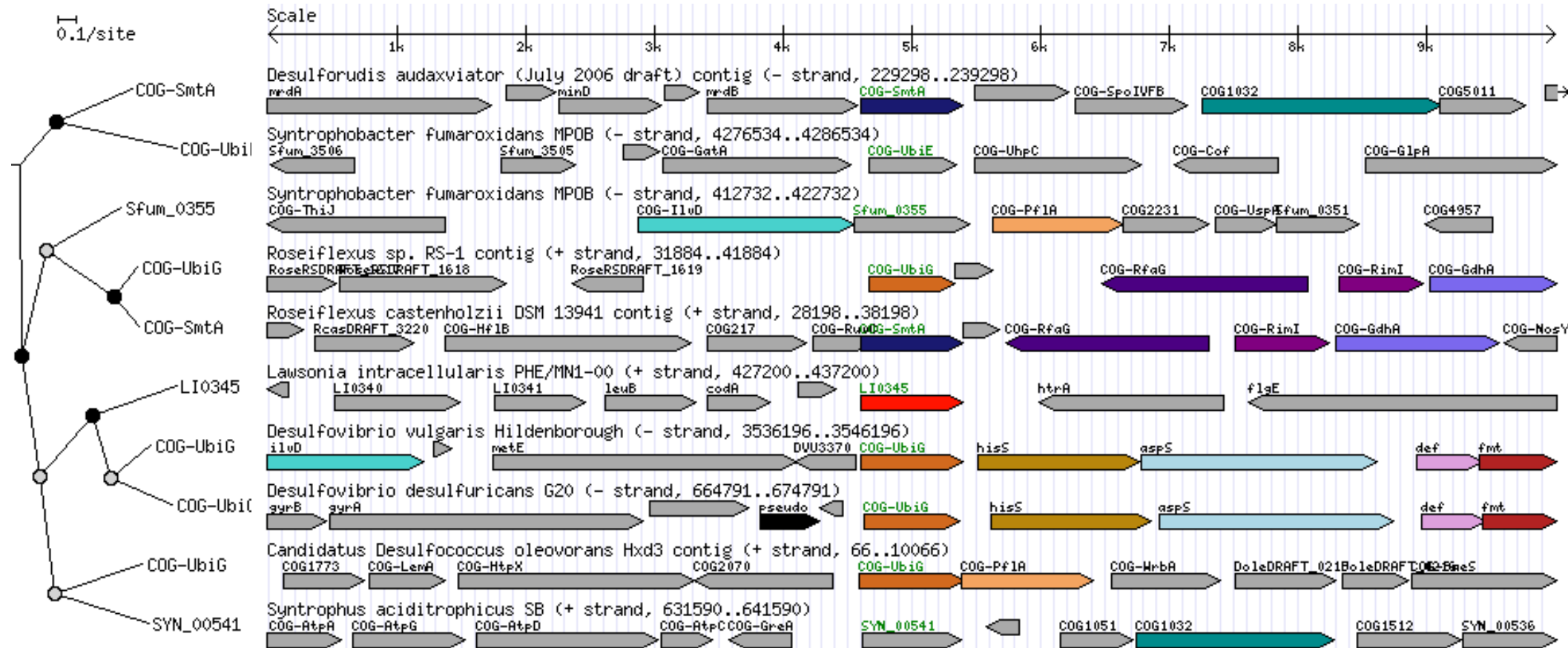
Activation & Sulfhydrylation



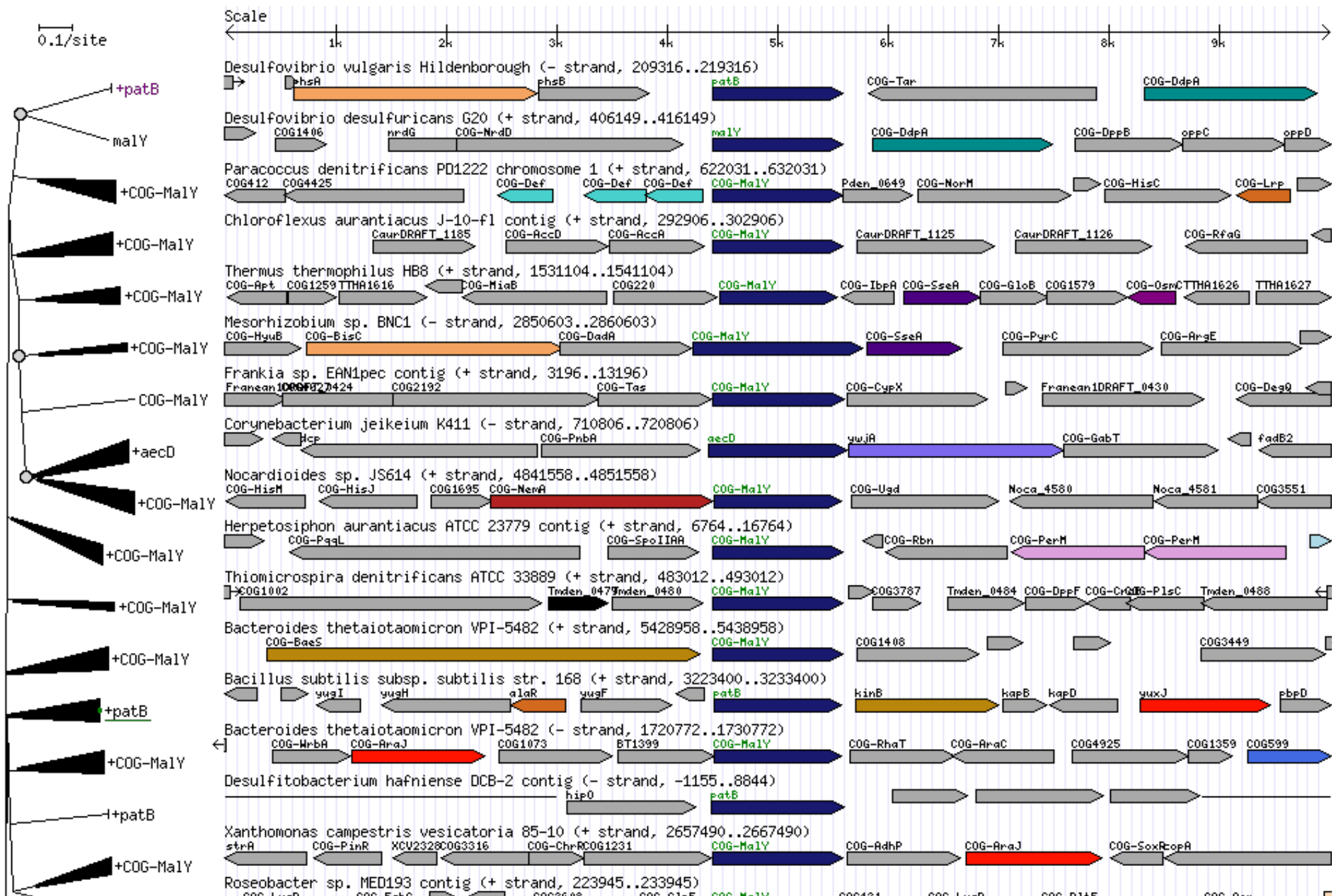
DVU3369, MetW?

Description	Domain ID	Range	E-value
Magnesium protoporphyrin O-methyltransferase, bacterial	PIRSF036745		0.0015
UbiE/COQ5 methyltransferase	PF01209		0.00048
HEXAPRENYLDIHYDROXYBENZOATE METHYLTRANSFERASE	PTHR23134		1.5e-06
2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol methylase	COG2227		4e-12
S-adenosyl-L-methionine-dependent methyltransferase	SSF53335		2.7e-30
Methionine biosynthesis MetW	PF07021		0.0034
METHYLTRANSFERASE	PTHR10108		2e-11
Vaccinia Virus protein VP39	G3DSA:3.40.50.150		1.6e-25
METHYLTRANSFERASE-RELATED	PTHR12734		2.1e-06
FAMILY NOT NAMED	PTHR16458		0.0059
Methyltransferase type 11	PF08241		9.1e-24
Methyltransferase type 12	PF08242		4.7e-16
[low-complexity (repetitive) sequence]	seg		--

DVU3369, MetW?



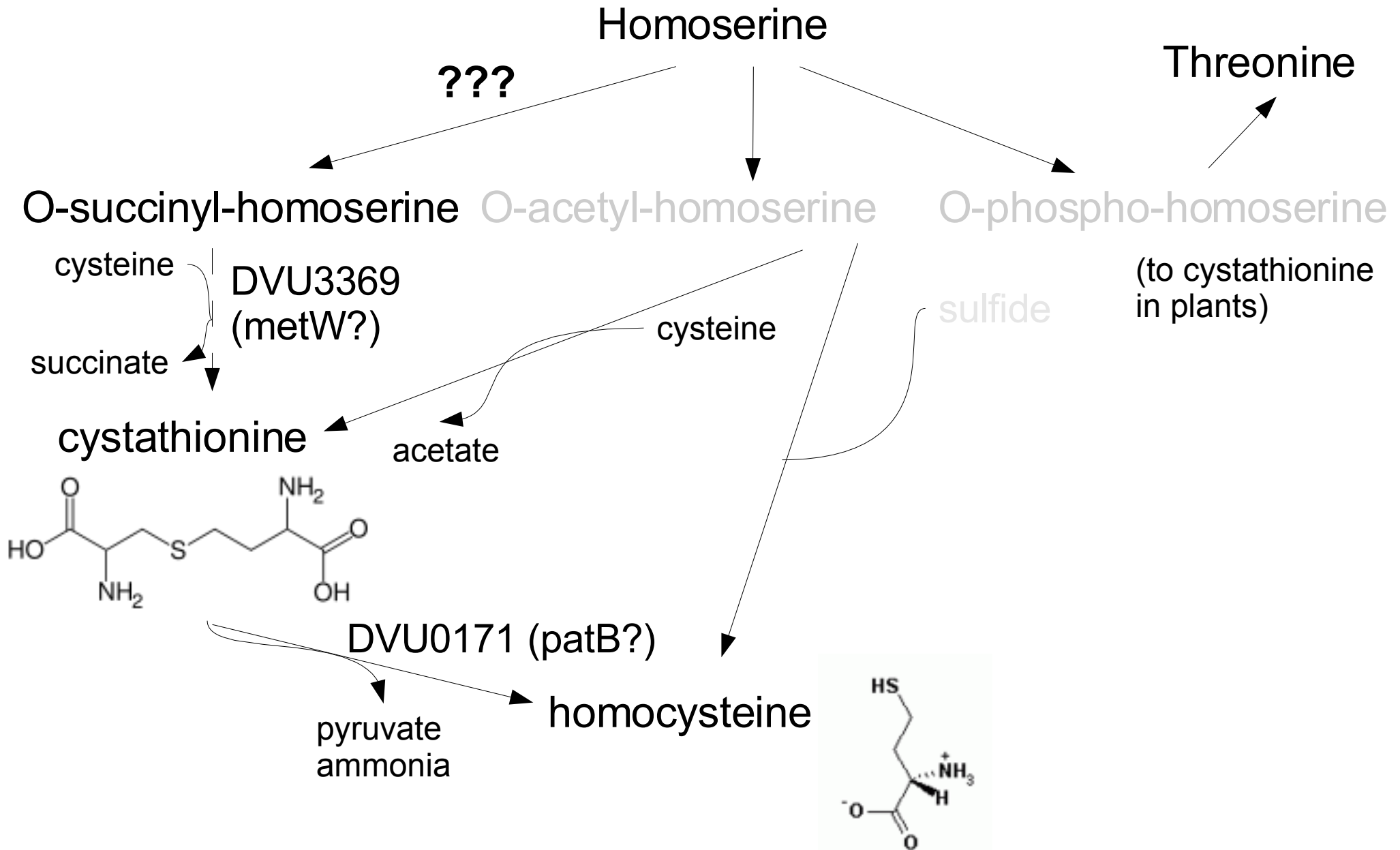
Conserved in G20, *Lawsonia intracellularis* (1185 genes)
 Near MetE (homocysteine methyltransferase) in DvH
 Constitutively expressed, growth-rate-dependent

[illegible]

DVU0171, Cystathionine beta-lyase?

- Similar to *B. subtilis* patB
 - major function is as cysteine desulfhydrase
 - also a cystathionine beta-lyase *in vitro* and *in vivo*
- and *C. glutamicum* aecD
 - Beta-C-S lyase, can cleave cystathionine
- More speculatively, DVU0171 could transfer sulfur from cysteine to homoserine (via cystathionine)

Proposed Methionine Synthesis Pathway



Questions about Methionine Synthesis?

Rapid Evolution of Gene Regulation

- Goal: Predict gene regulation in DvH
 - Interpretation of expression data
 - Aided by expression data
- Annotate transcription factors (TFs)
 - e.g., PerR, Fur, RpoH
 - but most cannot be annotated
- Predict binding sites
 - e.g. conserved upstreams in DvH and G20
 - but most are not conserved for so long

How Gene Regulation Evolves

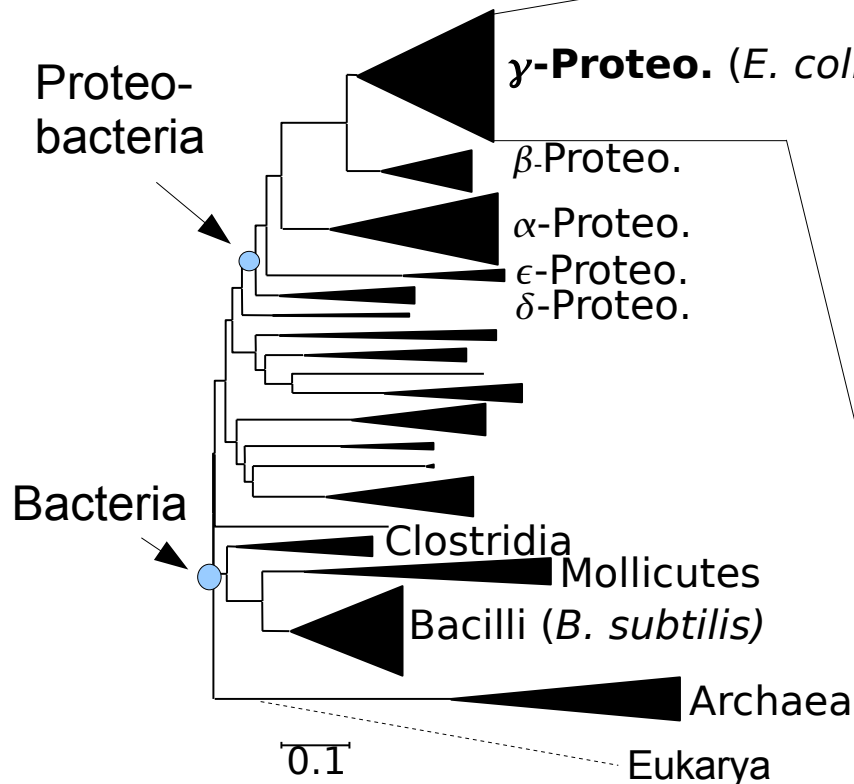
- Studies of model systems (*E. coli* & *Shewanella*)
 - Histories of TFs
 - Histories of interactions

Histories of TFs

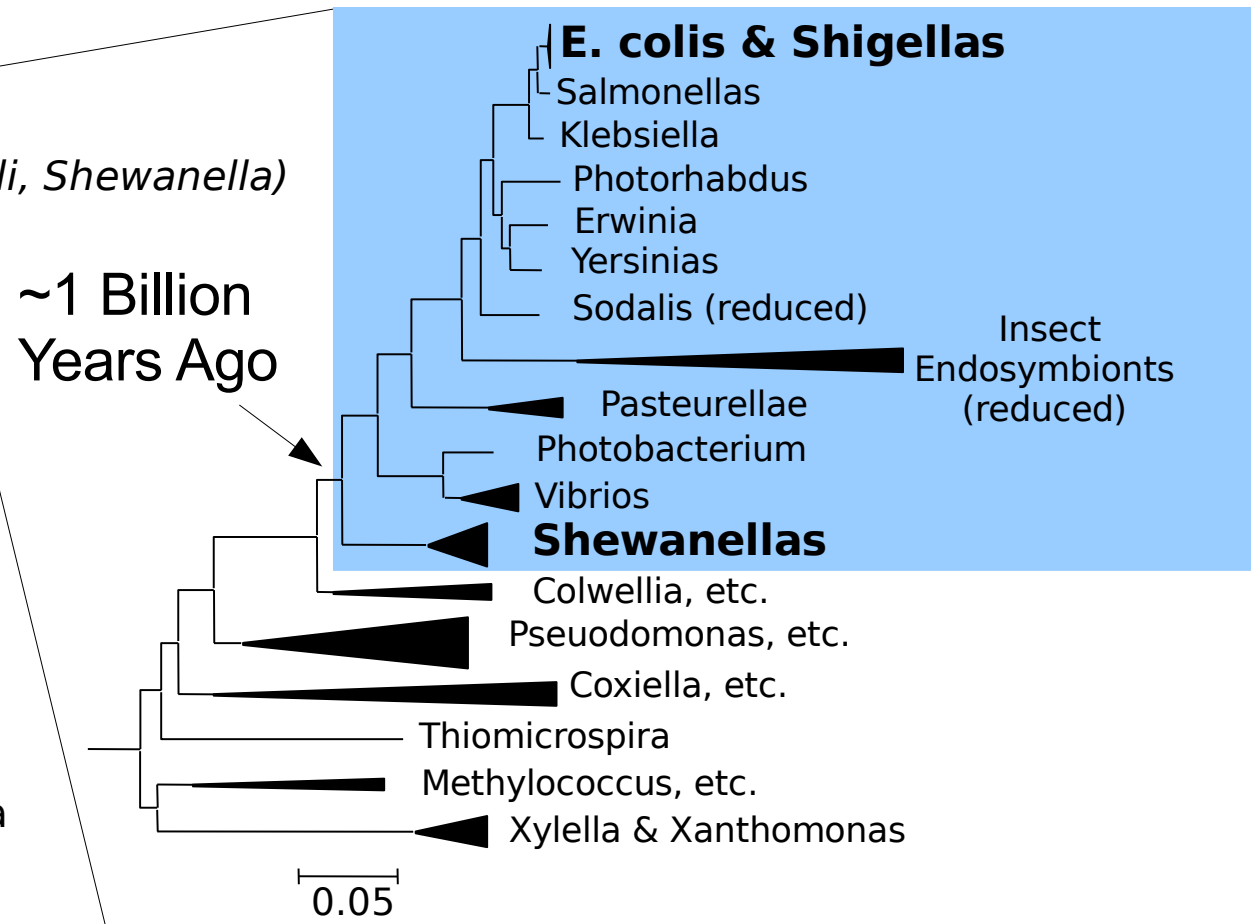
- Reconcile gene tree to species tree
 - Congruence: vertical inheritance
 - Horizontal gene transfer (HGT)
 - Duplication

Species Tree

The Tree of Life



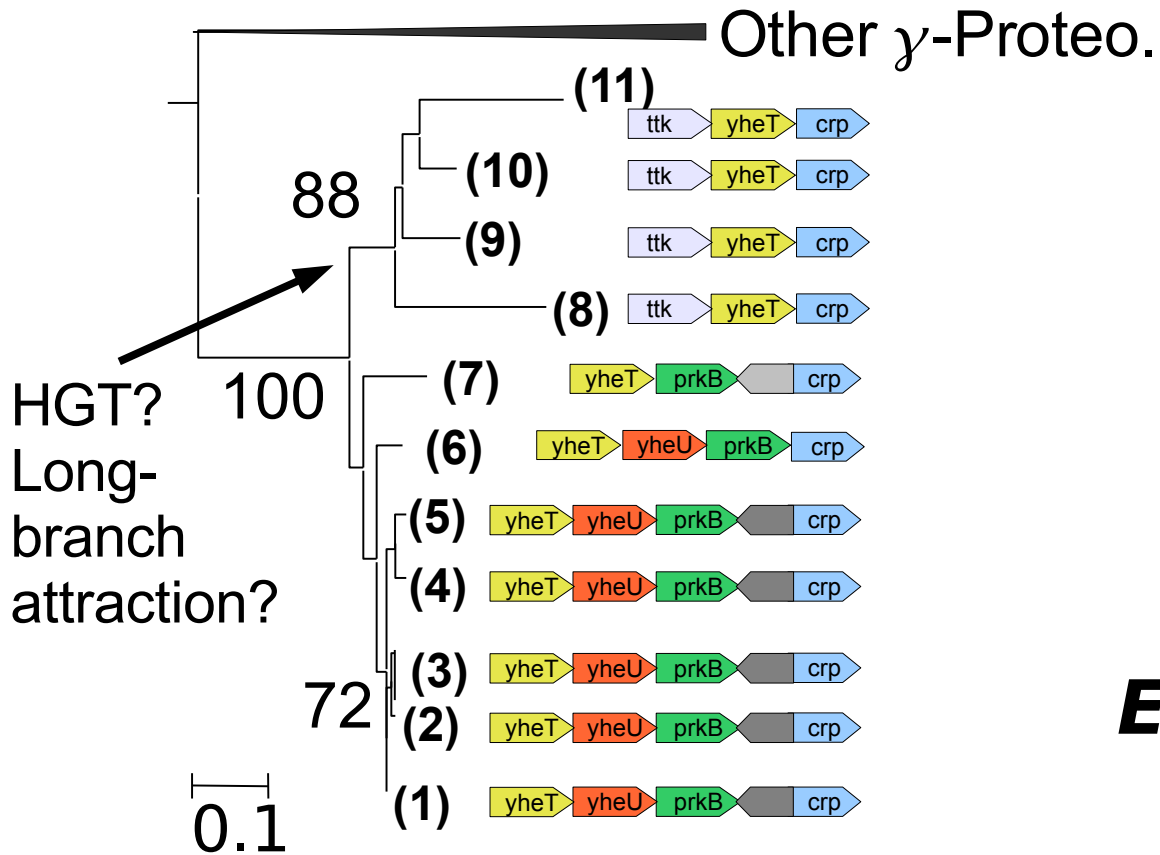
The γ -Proteobacteria



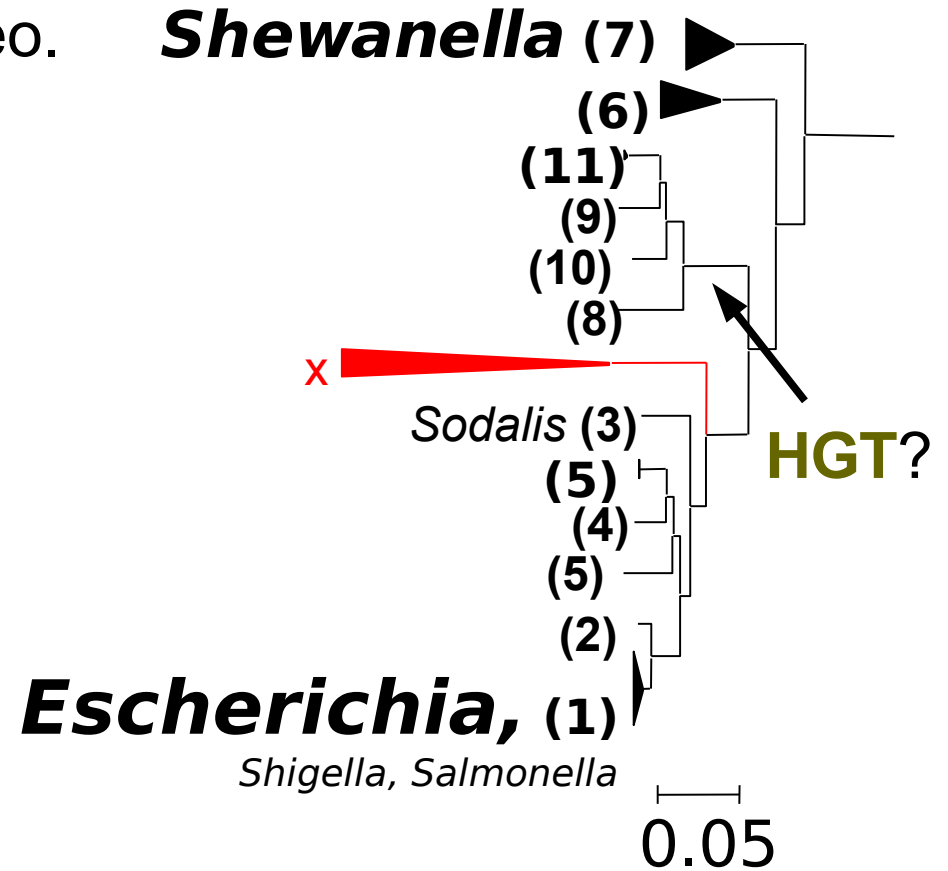
~ 1 billion yrs. since divergence of *E. coli* & *Shewanella*
Comparable to DvH/G20 distance

Vertical Inheritance of Global Regulators

Gene tree for crp



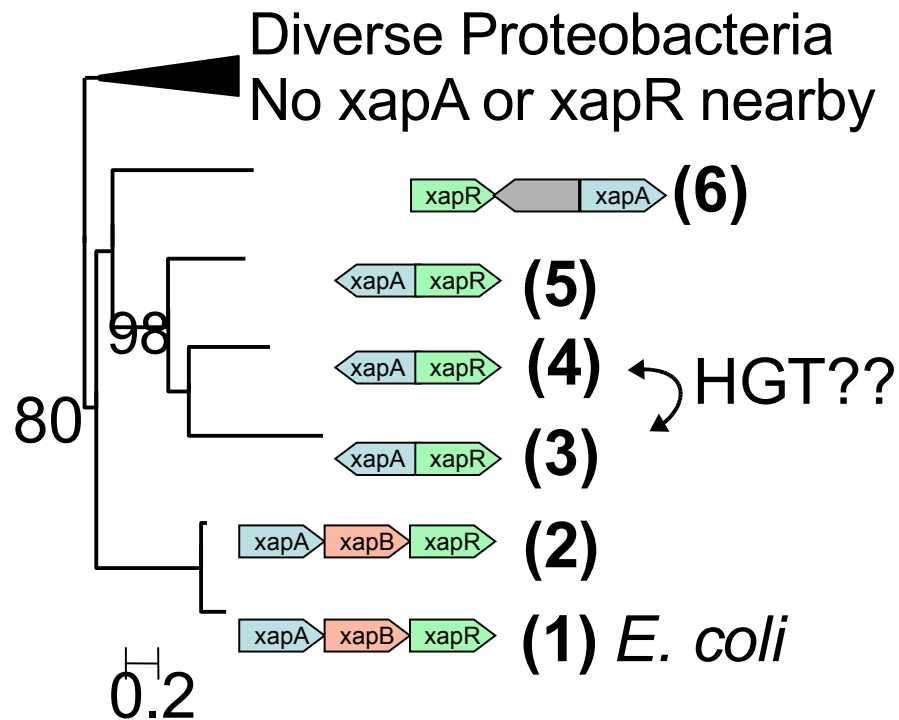
Distribution of crp



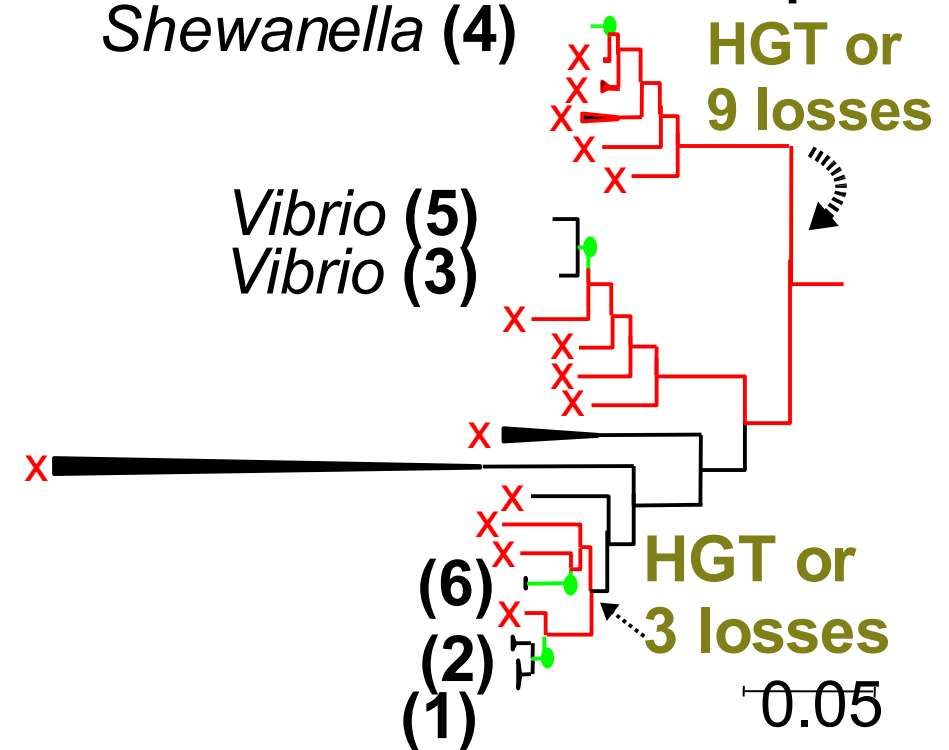
17/20 top global regulators are native

Complex Histories of Neighbor Regulators

Gene tree for xapR



Distribution of xapR

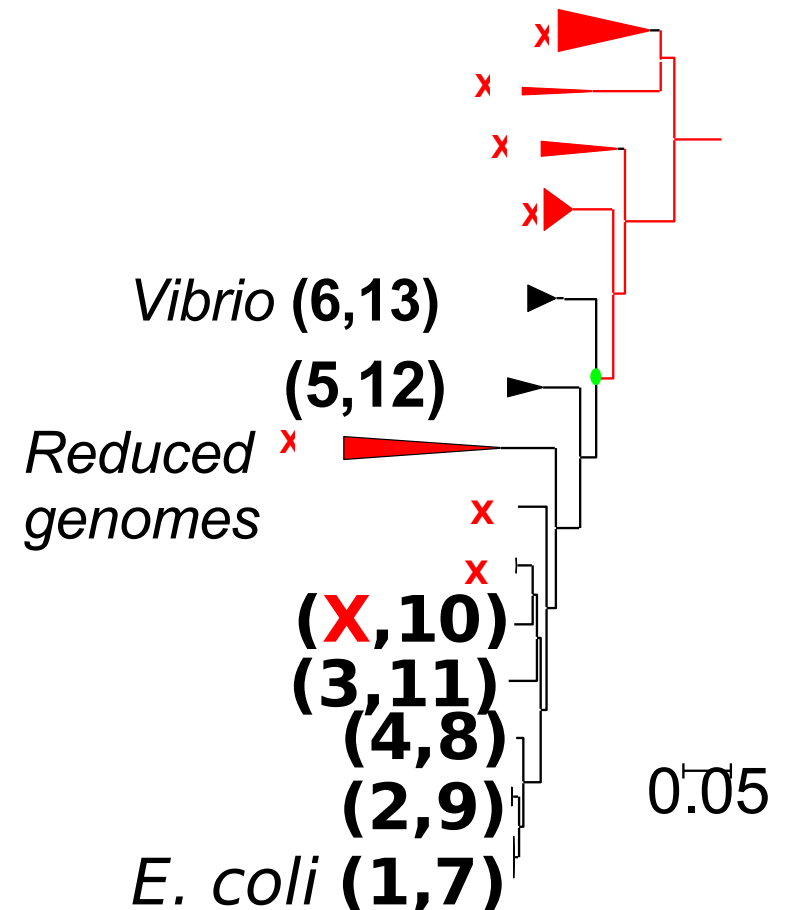
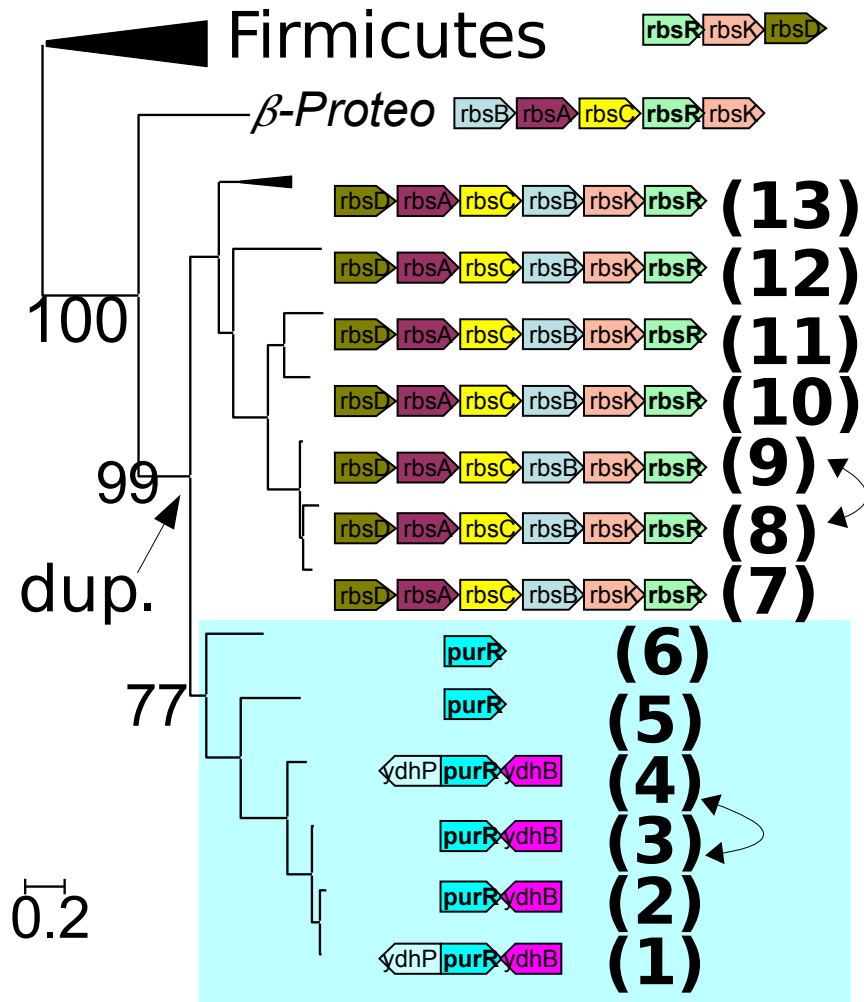


- Co-transfer: ~60% of neighbor regulators, 45% of putative regulators => predictions
- Repeated HGT: ~40% of neighbor reg.

Duplication of rbsR/purR

Gene tree for rbsR, purR

Distribution of rbsR, purR



- Dups are rare (13% of TFs)
- Non-overlapping functions (~half of dups)

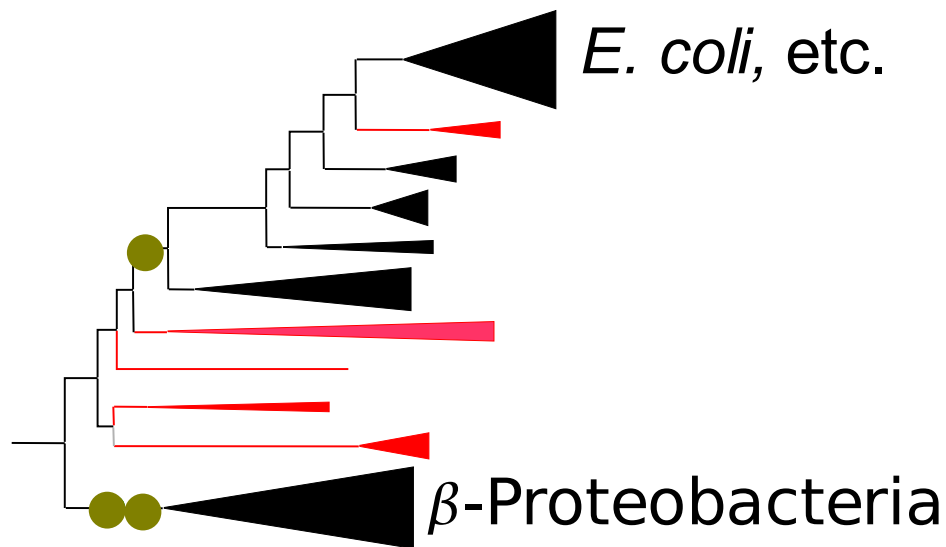
What About Older Histories?

- Use bidirectional best BLAST hits (BBHs)
 - often used for annotation
 - few TFs have BBHs (~25%, *E. coli* to *B. subtilis*)
- Global regulators not in distant bacteria
 - Mostly not in *B. subtilis* (4/20 present) or DvH (6/20)
- Do regulators have conserved functions, even if history is complicated?

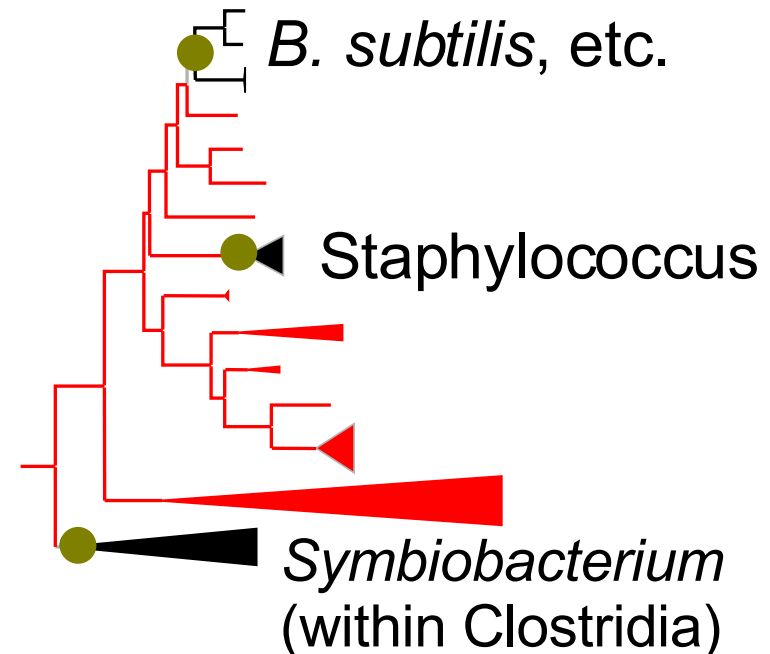
BBHs of TFs Have Different Functions

- Different pathways & stimuli
 - *E. coli* *betI*: choline → osmotic stress
 - *B. subtilis* *pksA*: polyketide synthase

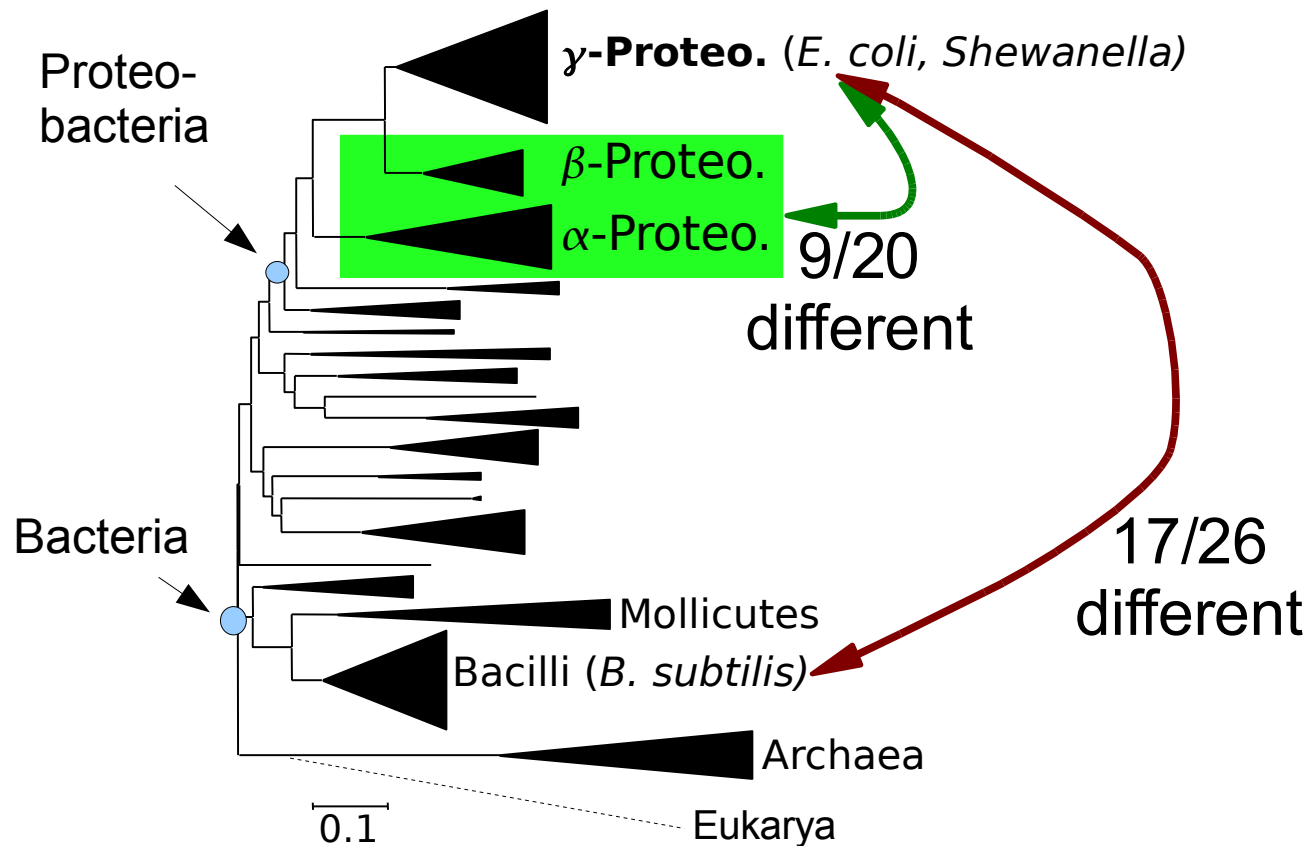
***betI*: HGT or 3 losses**



***pksA*: 2 HGT or >5 losses**

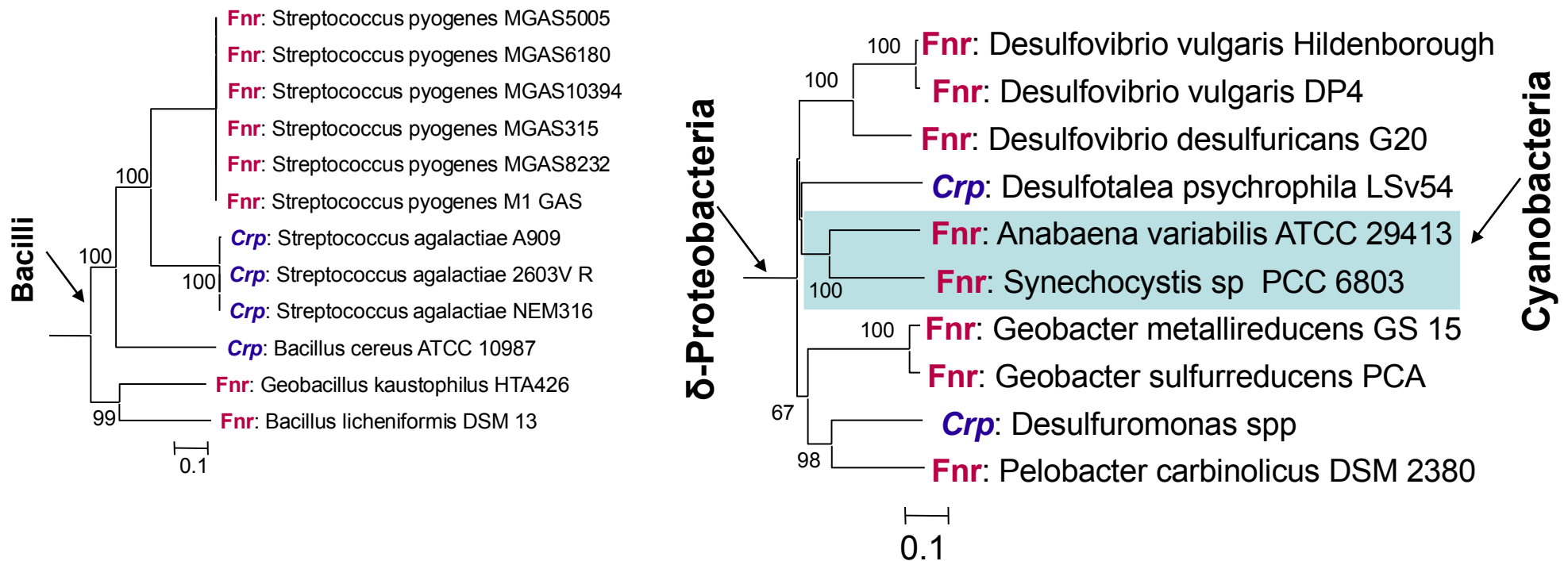


Characterized BBHs of TFs Have Different Functions



BBHs with different functions have complex histories (duplication or >1 HGT event)

“Orthologs” of **Fnr** and **Crp**



DvH/G20 “Fnr” is predicted to be hcpR
senses nitric oxide or nitrate, not oxygen
binds CRP-like sequence TTGTGA, not TTGAT

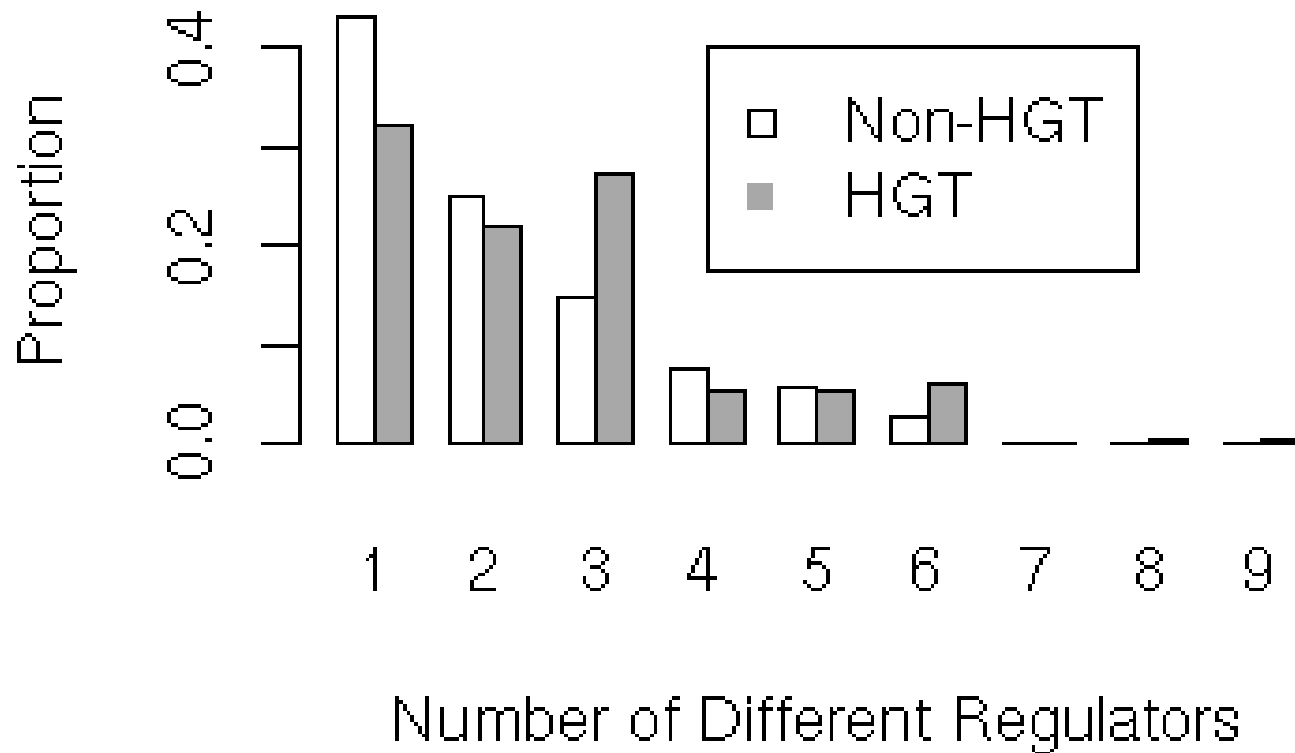
Summary of TF Histories

- The Last Billion Years
 - Little duplication (13%)
 - Lots of HGT (63%)
 - Complex HGT of neighbor regulators vs. conservation of global regulators
 - Co-transfer => predictions
- Distant bacteria
 - homologs (even BBHs) have different functions
 - most TF annotations are wrong

Histories of Regulatory Interactions

- TFs evolve “rapidly”, but 2/3rd of regulation in *E. coli* is by top-20 regulators
 - do their regulons evolve quickly?
- How HGT genes regulated?
- Evolution by duplication?

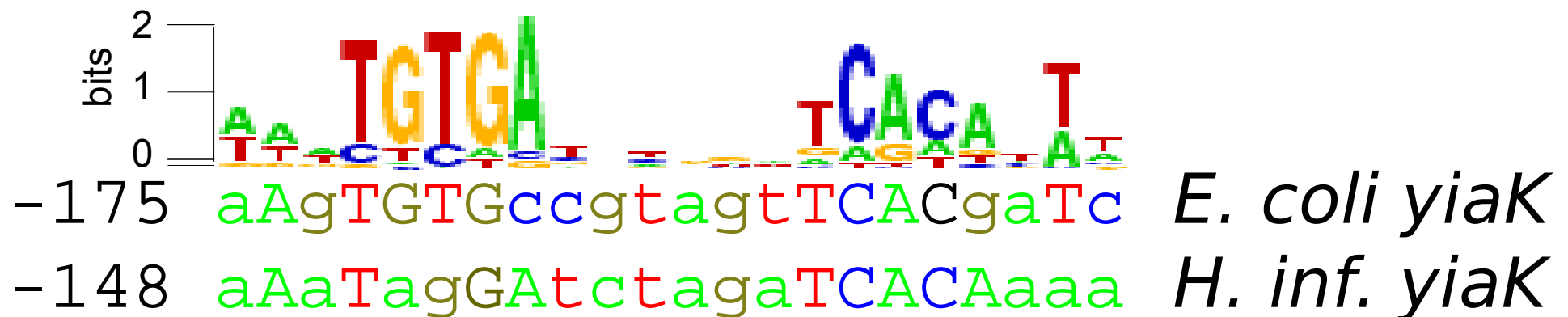
Complex Regulation of HGT Genes



E.g., CRP regulates half of HGT genes but only a quarter of other genes

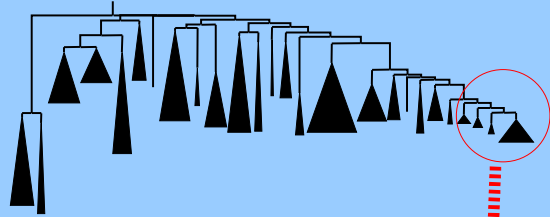
Sources of the CRP Regulon

- From distant bacteria w/o CRP (~80%)
- From related bacteria with CRP (~20%)
 - CRP site conserved across HGT in 4/12
- Sites usually not conserved across HGT (6/20 for global regulators)
 - except for co-transfer (presumably)

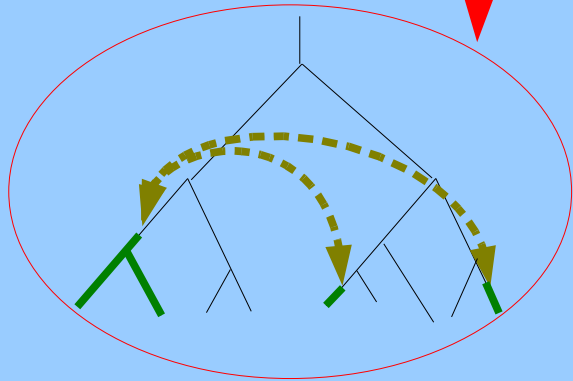


Niche-Specific Neighbor Regulators vs. Conserved Global Regulators

Tree of Life



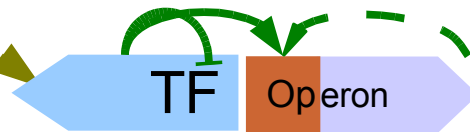
Last 1 Billion Years



Repeated transfers
between related
bacteria

Co-transfer w/ operon

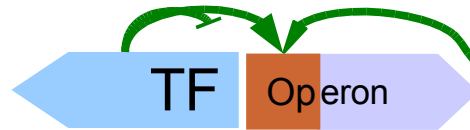
HGT



*Conserved
Global
Regulator*

~~GR~~

and gain a new site

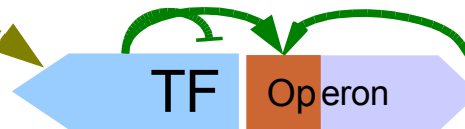


GR

Time ↓

and transfer again with the site

HGT

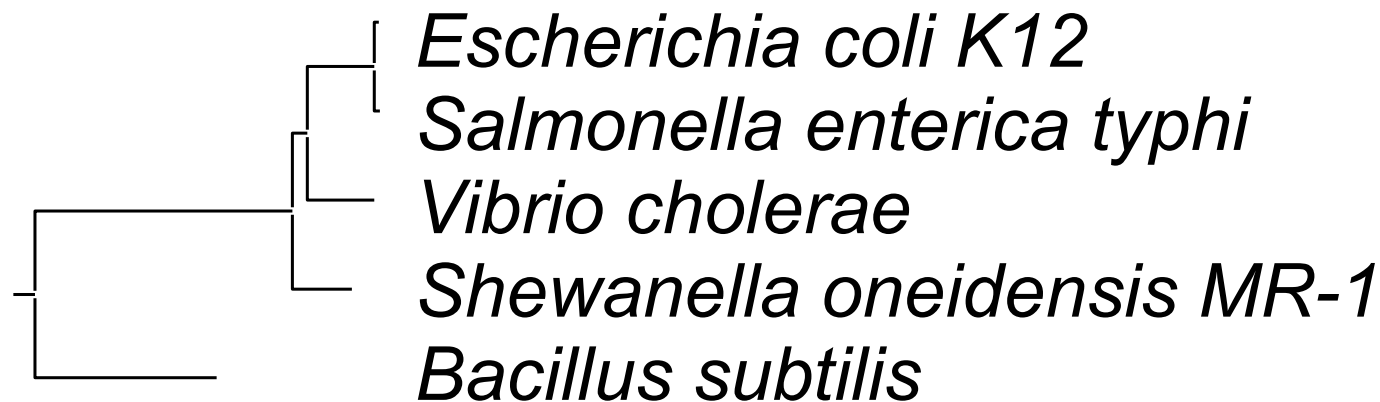


GR

(not HGT)

Conservation of Regulation

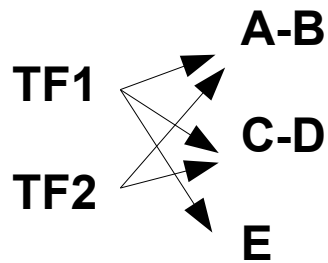
- Predict that regulatory interactions are conserved (if BBHs are present)
- See if coexpression is conserved



0.05

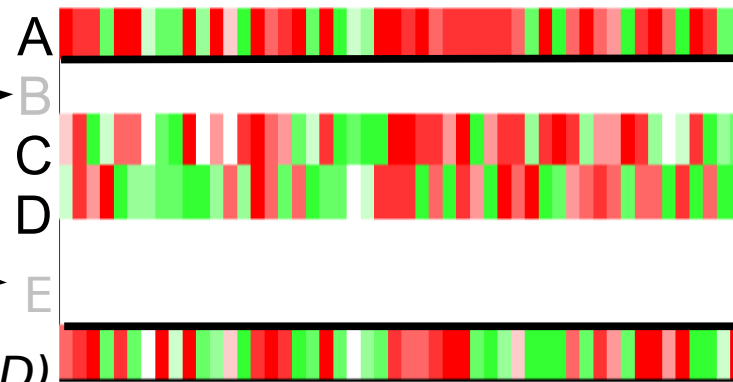
Gene-Regulon Correlations

Regulatory Network
in *E. coli*



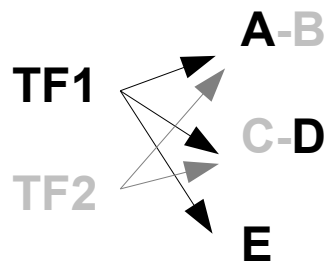
same
operon →
different set
of regulators →
 $average(C,D)$

Gene-Regulon Correlation
for Gene A in *E. coli*



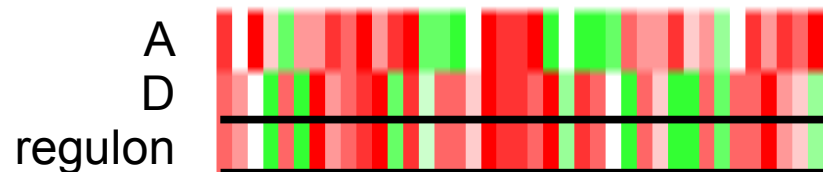
$$\text{Gene-regulon cor}(A) \text{ in } E. coli = \text{cor}(A, \text{average}(C,D))$$

Putative Regulatory
Network in Relative



Absent genes: TF2, B, C

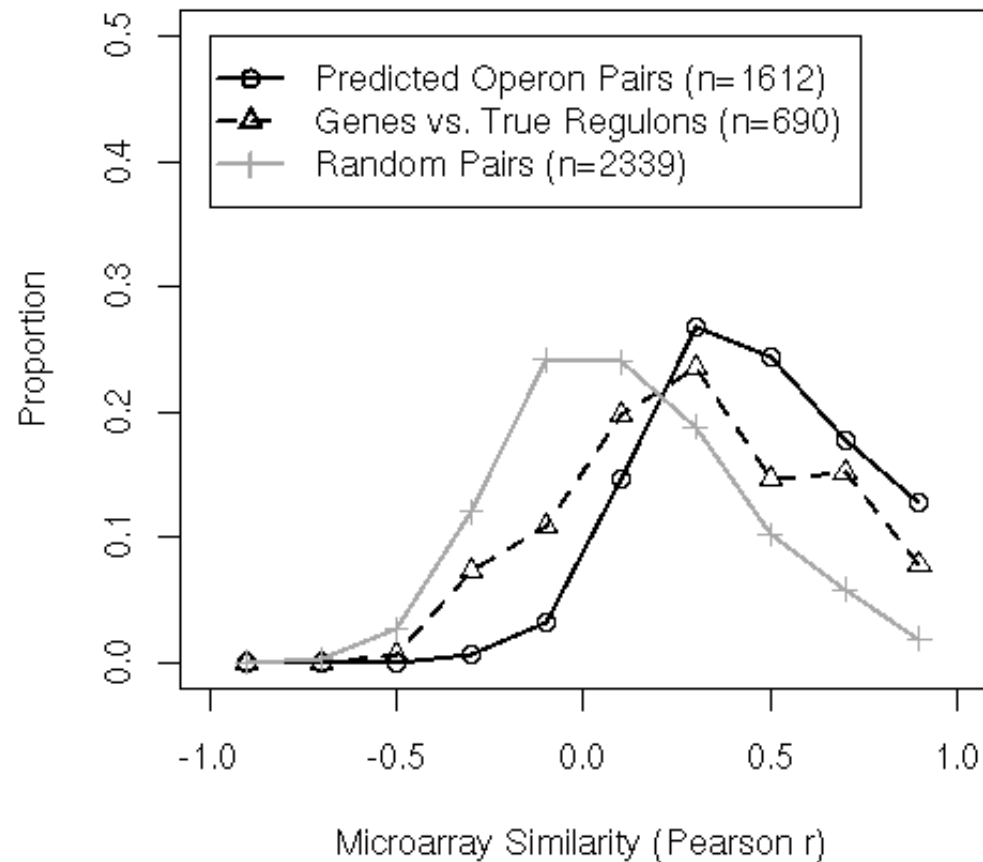
Gene-Regulon Correlation
for Gene A in Relative



$$\text{Gene-regulon cor}(A) \text{ in Relative} = \text{cor}(A,D)$$

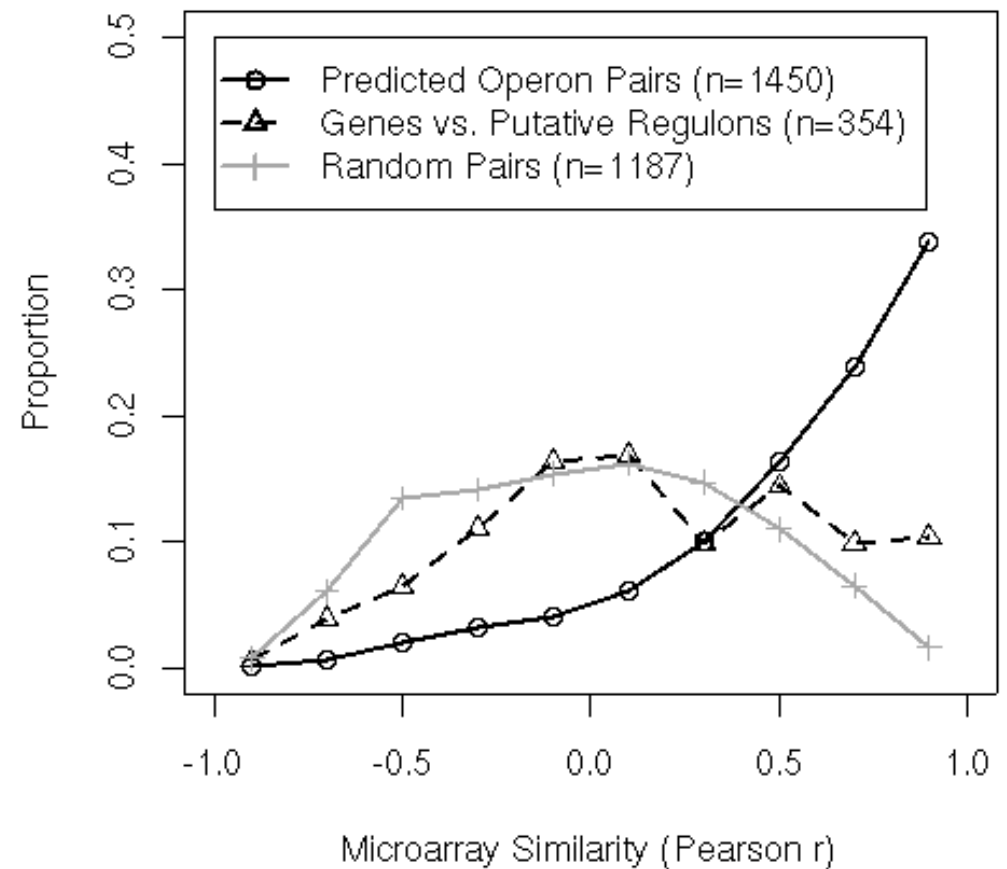
Gene-Regulon Correlations

(A) *Escherichia coli* K12



79% as coexpressed as operons

(D) *Shewanella oneidensis* MR-1



55% as coexpressed as operons

Controls for Gene-Regulon Correlations

How Coexpressed?

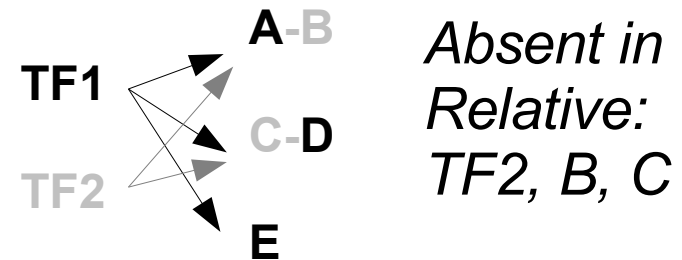
Operons → perfect coexpression

Random genes → noise

$$\text{Coexpression ratio} = \frac{\text{mean(regulons)} - \text{mean(noise)}}{\text{mean(operons)} - \text{mean(noise)}}$$

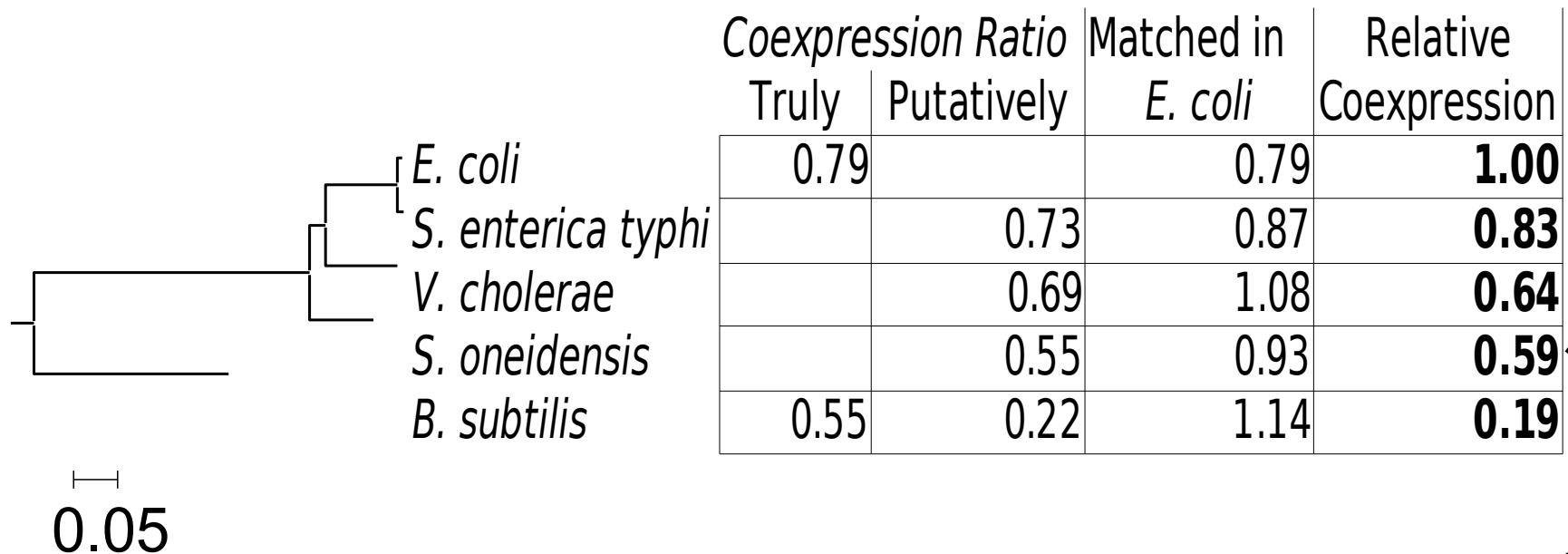
How Accurate?

$$\text{Relative coexpression} = \frac{\text{coexpression ratio in Relative}}{\text{matched ratio in Ec}}$$



Matched-in-*E. coli*
gene-regulon $\text{cor}(A) = \text{cor}(A,D)$

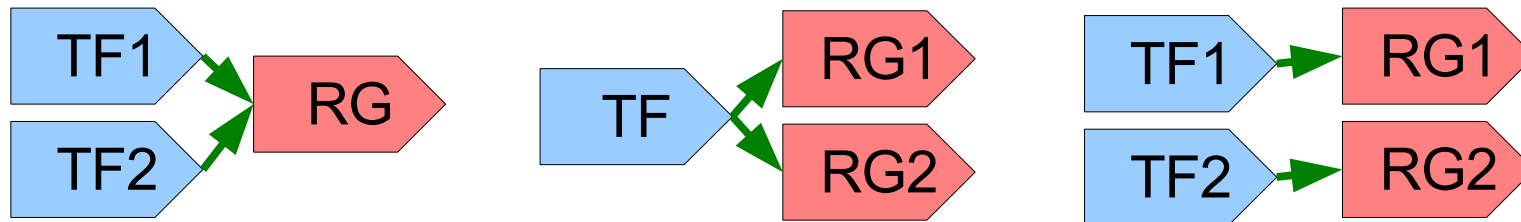
Limited Conservation of Regulation for Conserved Genes



- From shuffling the *E. coli* network,
59% coexpression = ~70% conservation
 - modest conservation to *Shewanella*
 - validated changes to *arcA*, *fnr*, *fliA* regulons

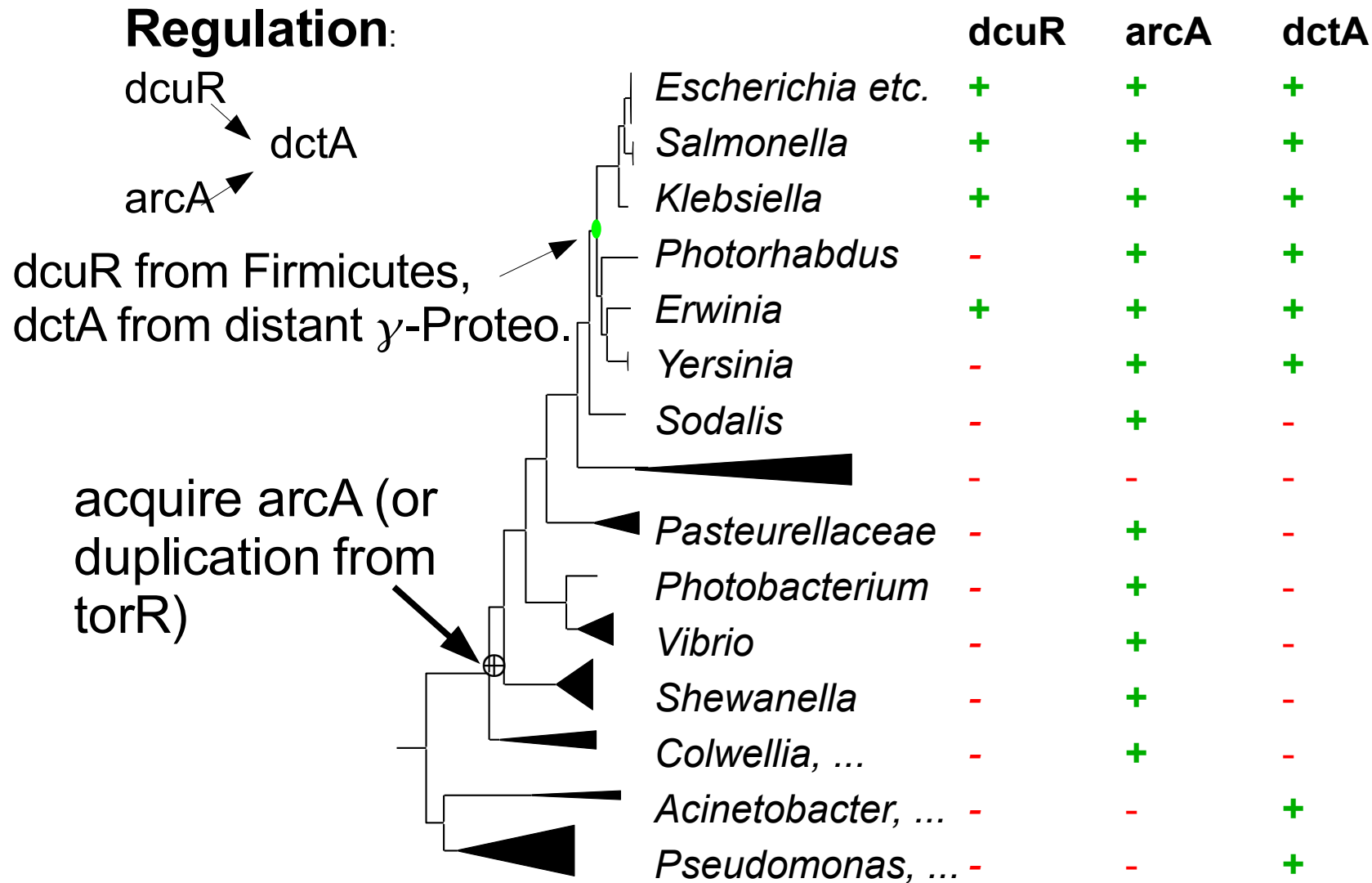
Evolution of Regulatory Interactions by Duplication

- Three Scenarios



- More common than chance (Teichmann & Babu 2004)
 - analyzed distant paralogs
 - is it really conserved from common ancestor?

Age of Regulation vs. Age of Duplication



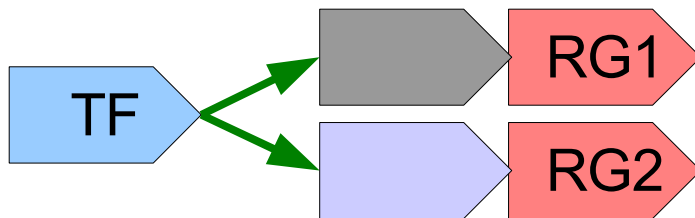
Convergent Evolution

- Not conserved from common ancestor
 - Regulation cannot be older than presence of RG & TF in the lineage
- Regulation of dctR evolved after transfer
- Analyze shared regulation for distant paralogs
 - 29/30 cases are convergent evolution

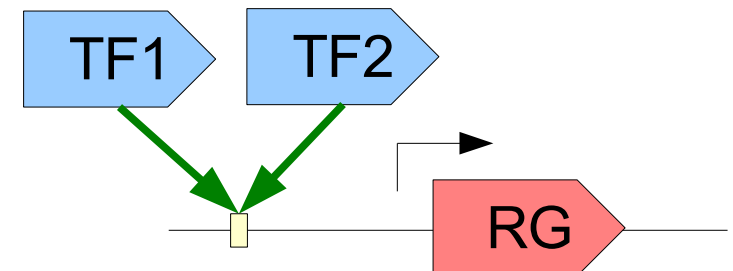
Convergent Evolution of Regulation

- More recent paralogs (>30% identical)
- Analyzed 425 shared TF-RG interactions
 - 14% of RegulonDB
 - 5% duplication
 - 6% convergent evolution (>chance, $P < 0.001$)
 - 3% unclear

Operon structure inconsistent with evo. by duplication (166)

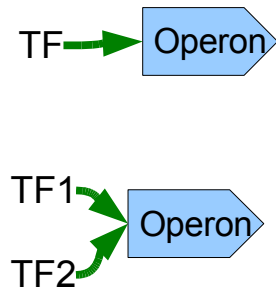


Convergent evo. of a single shared site (28)

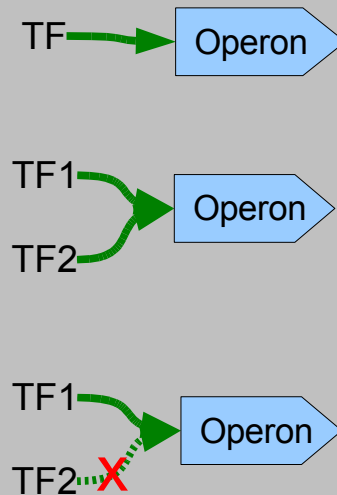


Rapid & Convergent Evolution of Regulatory Interactions

Duplication & conservation

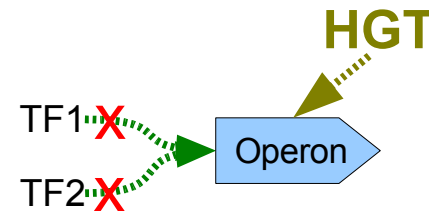


Duplication & divergence

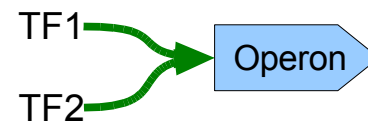


Duplication & convergence

Acquire an operon

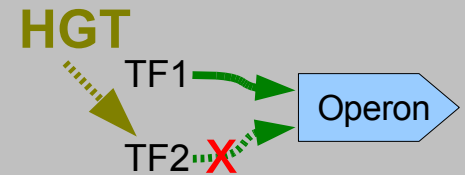


Evolve shared site
(similar DNA binding)

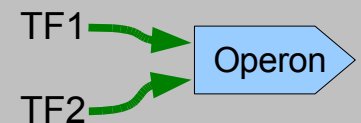


HGT & convergence

Acquire a paralog



Evolve new site
(not shared)



Rapid Evolution of Gene Regulation

- Transcription factors (TF) have complex histories
 - Rampant transfer (not duplication)
 - “Orthologs” are problematic
 - Annotations are usually incorrect
 - “Neighbor regulation”
 - driven by horizontal gene transfer, aids annotation
 - Global regulators are more conserved
 - but we can predict little about *D. vulgaris*
- Regulatory interactions not highly conserved
 - e.g. *E. coli* to *Shewanella*, or *D. vulgaris* to G20